# Digitization and Archival Information

### Diane Vogt-O'Connor

Every decade or so a new technology appears on the information horizon that is heralded as the replacement for all audiovisual and textual documents. In the 1950s it was microfilm. In the 1970s it was videotape. In the 1980s it was videodiscs; while in the 1990s it is CD-ROMs and digitization that will lead to paperless offices and archives and bookless libraries. Archives and libraries will vanish in favor of home information centers! Access to all information will be instantaneous! Archives, libraries, and museums can throw out their originals!

Humans are naturally resistant to new technologies. Being tactile creatures we like information sources we can touch, manipulate, stroke, and hold in our hands. We like to be able to sit in our easy chairs and browse conveniently packaged information while we eat or drink. We read in the bathtub. We like a profoundly personal relationship with our information sources. IBM recognized this fact when they called their micro-computers "personal computers."

Our resistance to technology is enhanced when it becomes clear that new technologies are rarely good for the purposes for which they are planned and sold. When the personal computer was first marketed we were told that it would balance our checkbooks and tell us what to make for dinner and how (i.e., provide recipes) based upon the contents of our refrigerators. What we actually use our personal computers for is, by and large, wordprocessing.

### Digitization and Museum Archives

CD-ROM and digitization are new technologies that have yet to prove what they can do most effectively. (Note: CD-ROM means compact disk read-only memory. Read-only refers to the fact that the data on the disk may not be changed by the reader.) We are told that CD-ROMs and digitization are the solution to all of mankind's informational woes. Caution is advised. So far, their most effective products have been games and portable libraries, not access to original documentary sources. This may change if most archives, libraries and personal computer users can be convinced to invest in CD-ROM drives. On the other hand, another new technology may pop up momentarily that is incompatible with CD-ROMs. Museums and archives are generally impoverished organizations that can not afford to invest in transitory systems.

Currently, the actual market for archival documentation on CD-ROM is fairly small and relatively few archives or libraries are producing CD-ROMs. To be effective, CD-ROMs must be operated with a database package to provide searchability. Since most collections in archives and libraries, particularly photographs, are not consistently labeled, or identified it is very labor intensive to build effective databases for them. Such database building involves research, data verification, name and term standardization, input, and editing. Without this work, the major advantage of CD-ROMs, easy searchability, is lost. Some of these problems can be solved for textual data, by the use of full-text searching, but not all.

### New Technology Poses New Problems

Most new technologies lose more than they gain in the process of duplicating archival and manuscript materials. Microfilm drove researchers crazy due to the awkward reader/printer access systems, poor indexing, and poor quality control (frequent illegibility). Videodiscs were expensive, required special equipment to play, impossible to update, and often had mediocre image quality. CD-ROMs are short-lived and require special equipment to play.

Perhaps most worrisome, digital data requires a high original data capture investment; an endless continuing investment in data migration to each new generation of hardware and software; and a need for data verification and certification after each transfer. Once digitized, data is more vulnerable to destruction. Digital data is short-lived in most of its forms. Diskettes are fragile. Computer tape and videotape last between 10 and 20 years if played and checked annually. While CD-ROM recording media may last 100 years, the disks themselves delaminate after between 10 and 15 years (depending upon which conservator or research scientist is speaking), leaving users with a useless pile of plastic.

### Digitization: Legal Problems

Beyond these issues, digital data, particularly CD-ROMs, frequently run aground on the shoals of copyright and privacy legislation. Archives, museums, and libraries frequently do not have the copyright to the images and manuscripts within their collections. Instead, the copyright belongs to the creator (author or photographer). Reproducing these manuscripts or images on CD-ROM or the Internet is publication and thus a violation of copyright.

Privacy legislation is even more complex. Living individuals have a right to privacy. Publishing their face or words without permission is illegal. Therefore, placing the oral histories, transcripts, or images of private living individuals on a CD-ROM, the Internet, or other distributed digital source is illegal without prior written permission from the individual illustrated or quoted. Significant portions of most park audiovisual collections show unnamed individuals, just as many park collections of original audiotapes, motion pictures, photographs, and videotapes lack release forms.

### Digitization and Original Documentation: What is Lost

Finally and perhaps most conclusively, digital data only captures a portion of the functional values of

archival documents such as diaries, letters, photographs, and videotapes. These values include the following:

- artifactual value (value as a piece of material culture in a particular process and/or format and by a particular creator);
- associational value (value due to being owned, created, used, or evaluated by a particular individual or group);
- evidential value (legal value as unmanipulated evidence of an activity, event, function, or entity such as the answer to the classic question, "what did he know and when did he know it?");
- administrative value (as a record of the functions, policies, procedures, financial, and legal status of an organization); and informational value (recording the who, what, where, why, and when that journalists crave including names, dates, individual's opinions, activities, events, processes, and memberships.)

None of the new technologies can reproduce these values effectively. Consider CD-ROMs attempting to reproduce photographs. Photographic historians, scientists, and students of material culture are interested in the artifactual value of photographs. Knowing a photograph's process (e.g., cellulose nitrate negative or gum bichromate print) and format (cabinet card or mammoth plate) provides valuable information that helps researchers date, attribute, and determine what information could be recorded by the camera. Knowing the process also tells us how the information will degrade over time.

In the NPS, less than 1% of the existing photographs are labeled with this information. This photographic process/format information can only be obtained by examining the original image's tonality, paper fibers, deterioration, and the actual layer structure of the emulsion, binder, and base configuration. Scholars working with CD-ROMs will not be able to tell what process the image is, thus losing valuable information.

Digitized images are very easily manipulatable; thus, they are highly suspect as evidence. Photography, whether justly or not, has had a reputation for veracity. This reputation has led to photographs being used as evidence in legal cases, publications, and exhibits. As Lewis Hine noted, "Photographs don't lie, but liars photograph." Due to this reputation, photographs have been used to prove the existence of the non-existent from 19th-century photographs of pixies and ghosts to 20th-century images of the Loch Ness monster. When the new technology of digitization tries these tricks, however, it leads to an uproar. When *National Geographic* magazine chose to digitally rearrange the Egyptian pyramids in a cover image, the bad publicity was phenomenal. Digital images do not enjoy the same reputation for veracity as photographs either in the popular press or the courts.

By their nature, digital images do not have associational value. Recently, the NPS turned down an offer from the Library of Congress to receive a facsimile of the *Gettysburg Address* instead of continuing a long-term loan to the park of the original. The original *Gettysburg Address* made Lincoln's response to the tragedy at Gettysburg more immediate and poignant for park visitors as the document was directly linked to the event, as well as being written in Lincoln's own hand. A photographic copy does not have the same effect on visitors. Digital images are similarly mute where associational value is concerned. The hand of the master has vanished.

The forte of digitized documentation is information. Like xerox copies, digitized data can provide a significant portion of the surface information on an image or document. We can discover what is illustrated or discussed, although we will not be able to analyze the paper or process. In general, we also cannot examine the original signature or watermark unless the document takes up very large amounts of memory, thus becoming increasingly expensive.

If the materials are digitized intact, without cropping, in their original order, with care taken not to intermingle materials from different collections or sources, the resulting CD-ROM will provide significant assistance to scholars, particularly those who could not normally visit the archives or library. Thus, the CD-ROM has a democratizing effect, providing access to scholars without the money or inclination to travel or request copies. On textual documents, natural language full-text searching can mitigate the need for a database by providing limited accessibility to the more patient and innovative researchers who have a good dictionary of synonyms in hand.

## A Cost/Benefit Analysis of Digitization

It is, however, essential to consider the cost. Conservators do not recommend digitization as the initial cost of digitization and CD-ROM production is not significantly less than that of producing copy photographic negatives, while lasting about 1/20 as long as a photographic negative. This cost does not factor in the cost of transferring the data as the software and hardware change and as the magnetic media degrades every decade or two. Unlike photographs, digital data requires functional hardware and software that must be kept in good order and upgraded as necessary—another cost.

The benefits we receive for this increased cost are three: 1) increased searchability—if the research and database work is completed correctly (another significant cost); 2) better, more democratic distribution of data—if the copyright and privacy situation allows us to actually publish the CD-ROM; and 3) the capability to manipulate images and text by correcting the contrast, removing stains, and enhancing image sharpness—if we have the time and inclination.

The hidden costs of digitization are three: 1) potential copyright and privacy lawsuits; 2) the endless cost of data migration and new hardware and software as developed; and 3) the potential loss of the original item when money that might be spent on caring for the original item is diverted to a glamorous new technology.

The digital copies, which contain only a portion of the information found in the original document or photograph, are often produced as supposed replacements for the original. This is rarely done overtly. Instead, money that might have been spent on conserving the

originals or photographically or xerographically reproducing a larger portion of the information found in the originals is spent on digitization. This digitization is despite the fact that digital data is less durable, even when printed out in many cases. For example, printouts of digital data made in the dye sublimation process are so fragile and sensitive that they are proposed for use as pollution indicators. If not cared for, the original photograph or document deteriorates and we are left with a pale digital shadow that requires continual copying, verification, and migration to new software and hardware to keep usable. During times of rapid technological change, this is an excellent recipe for losing our informational heritage.

With recent improvements in material science abounding, this is of particular concern. As with rain forests, we are only now learning how to fully extract all the value from our original source materials. Who knows what new techniques we will have in the future for analyzing and studying our documentary heritage? If we allow the originals to self-destruct while funding digitization, we may destroy or lose an important portion of our heritage without ever realizing it.

### Keeping the Baby without the Bathwater

What is the solution? It is advisable for archives and museums— the impoverished custodians of the nation's heritage—to take a conservative approach to dynamic technologies, particularly in the time of cutbacks. First, we must take care of our original documents and images. The real document is not replaceable just yet by a digital copy. Maybe someday, but not now.

Second, analyze collections and begin systematically to gather standardized data necessary for describing and later distributing audiovisual and textual collections via databases and digitization. Finally, consider using outside funds to start electronic projects, so that baseline funding is not diverted from the care of collections. While awaiting funds, spend some time investigating the park's copyright and privacy situation.

Aldous Huxley called history "A branch of speculation, connected (often rather arbitrarily and uneasily) with certain facts about the past." Caring for our original archival and manuscript collections will ensure that less speculation needs to take place over the past of our sites, our parks, and the NPS.

––––––––––––––––

Diane Vogt-O'Connor is the Senior Archivist, National Park Service, Washington. She notes that she is very fond of the CD-ROM drives on her home and office computers, which she largely uses for reference purposes.

centers like the Western Archeological Conservation Center (WACC). This posed a dilemma: how to maintain an image collection for secondary uses and to reunite the images with their respective collections. It was felt that the real challenge to the photo archivist was not only to conserve the image collections and their project relationships but to create usable resources for future interpretive and comparative work.

### Turning Image Collections Into Image Assets

Once a project photograph collection has been properly accessioned, managed, and conserved, the archival rules for original order have been observed, and each image is associated with its respective project and descriptive documentation, two questions remain: where to store it and how to make the image available for secondary uses.

Part of the problem with the secondary use of these ordered photographic images is the need to re-sort, compare, and assemble images in ways that were not perceived when the image was first created. A second issue is the need to ensure the protection of the original image as a vital part of a project's documentation while maximizing the ability to reuse the image for secondary purposes.

The Southwest Region has initiated a program to create digital copies of its cultural and natural resource project images in order to resolve some of the issues that have been discussed above. It is the intent of the region to use these digital copies for most secondary uses—retiring the original images to project documentation repositories under the care of the Division of Curation or the National Archives.

There were six points that influenced the decision to create a synthetic digital image archive:

- The region did not wish to maintain an extensive photographic curatorial facility.
- Capturing a copy of the original slide or negative in a digital format was found to be substantially less expensive than making an interpositive or a duplicate slide.
- The digital image was easier to store and easier to find and retrieve for secondary usage than is the original.
- The existence of the digital image copy eliminated the need for ready accessibility to the original photograph and allowed the original to be curated at whatever facility is best suited to that task.
- The use of color film had become so widespread that it could no longer be neglected as "non archival." It was showing up in most collections that require archiving. This was especially true for slide collections associated with archeological projects.
- Creating multimedia assets from the photographic originals permitted the region to make far better use of its existing image collections.

It is clear that maintaining extensive photographic collections requires professional curatorial facilities and skills. Retarding the deterioration of some films requires carefully controlled conditions. Evaluation of and restoration of poorly conserved collections is the work of specialists. The region has only limited resources with